DOCUMENT RESUME

ED 432 176 HE 032 148

AUTHOR Porter, Stephen R.
TITLE The Robustness of the "Graduation Rate Performance" Indicators Used in the "U.S. News and World Report" College Rankings.
PUB DATE 1999-03-00
NOTE 36p.; Paper presented at the AIR-CASE Conference (2nd, Washington, DC, March 29-30, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Graduation; Higher Education; *Models; *Performance Factors; Predictor Variables; *Rating Scales; *Statistical Analysis; *Universities
IDENTIFIERS *AIR Forum; Graduation Rates; Performance Indicators; *U S News and World Report

ABSTRACT

This study examined the robustness of the "U.S. News and World Report" graduation rate performance indicator used in the magazine's college rankings. The graduation rate performance indicator is calculated as the difference between an institution's actual graduation rate and its predicted graduation rate, and is based on a linear regression equation controlling for student aptitude and institutional expenditures. Data were obtained for 198 of the 218 national universities used in the magazine's 1999 rankings. It was found that changes in the sample and variable definitions could cause the predicted graduation rate for an institution to fluctuate by plus or minus two percentage points. More refined model specifications reduced the number of institutions with extreme performance differences and could actually change the ranking of an institution from under-performance to over-performance, or vice versa. The use of confidence intervals for predicted graduation rates revealed that only about 5 percent of the institutions in the study had predicted graduation rates that differed significantly from their actual graduation rates. (Contains 34 references.) (MDM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 432 176

The Robustness of the "Graduation Rate Performance" Indicator Used in the *U.S. News and World Report* College Rankings

Paper presented at the 2nd AIR-CASE Conference
March 29-30, 1999
Georgetown University
Washington, DC

Stephen R. Porter, Ph.D.
Office of Institutional Studies
University of Maryland, College Park
College Park, MD 20742
Email: sporter@accmail.umd.edu
Phone: (301) 405-5608
Fax: (301) 314-9443

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Stephen Porter
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ABSTRACT

This paper analyzes the robustness of the *U.S. News* graduation rate performance indicator, calculated as the difference between an institution's actual graduation rate and their predicted graduation rate from a linear regression equation controlling for student aptitude and institutional expenditures. The sample is 198 of the 218 national universities used in their 1999 rankings.

Robustness is examined in four areas: the effect of small changes in the sample due to missing data or changes in how the sample of national universities is defined; the effect of seemingly irrelevant changes in variable definition; the effect of different model specifications that take into account additional measures of student quality and institutional constraints; and how the use of confidence intervals for the predicted values changes conclusions about performance.

Changes in the sample and variable definitions can cause the predicted graduation rate for an institution to fluctuate by plus or minus two percentage points. More refined model specifications reduce the number of institutions with extreme performance differences and can actually change an institution from under-performance to over-performance, or vice-versa. Finally, the use of confidence intervals for the predicted graduation rates reveals that only about 5% of the institutions in this study have a predicted graduation rate that significantly differs from their actual graduation rate. The implications of these findings for these types of models and recommendations for future research are discussed.

Introduction

U.S. News and World Report (USN) attempts to determine "America's best colleges" through their controversial annual college rankings[1]. These rankings are based on several different items, each attempting to measure some important aspect of institutional quality. One item, graduation rate performance (formerly referred to by USN as value added)

> … is designed to capture the effect of the college's programs and policies on the graduation rate of students after controlling for spending and student aptitude, which also affect graduation rates (Smith 1998 p. 35).

After regressing the *actual* six-year graduation rate for a new freshman cohort on both the average SAT score for the cohort and the amount of money spent by the institution per student, USN uses the statistical results to calculate a *predicted* six-year graduation rate for the institution. This rate is an estimate of what the institution's graduation rate should be given the quality of their students and institutional expenditures. The difference between the actual and predicted rates yields the performance indicator, so that

> If the actual graduation rate is higher than the predicted rate, *the college is enhancing the students' achievement* (emphasis added; Smith 1998 p. 35).

This indicator is an intuitively appealing input-output model: after controlling for student input (quality of the freshman cohort) and the constraints faced by the institution (the amount of money they are able to spend), one can easily determine what the output should be. If the actual output differs, we have some measure of how institutional policies such as faculty-student ratios, class size, etc., affect student behavior (graduation within six years). USN uses this approach to publish differences between expected and actual graduation rates for individual institutions, which they refer to as over- or under-performance. Figure 1 presents the distribution of the actual 1997 graduation rates by their predicted rates listed in the 1998 rankings for the USN sample of

---

[1] See for example Geraghty 1996, Machung 1998, Shea 1995, and Webster 1992.

national universities. The national universities appear quite varied in how well they graduate their students. Over one-quarter of the sample have predicted graduation rates that are ten percentage points higher or lower than their actual graduation rates (the maximum positive difference is +35 and the maximum negative is –19).

The USN indicator is simply one approach in a growing area of research attempting to assess institutional performance. Alexander Astin of the Higher Education Research Institute has advocated a similar methodology for estimating graduation rates (Astin 1997), as has the National Graduation Rate Study conducted within the American Association of Universities (Howard et. al. 1994, Kroc et. al. 1995, Kroc et. al. 1997) and the *Postsecondary Education Opportunity* newsletter (Mortenson 1997)[2]. These measures have undoubtedly been spurred in part by the growing demand in accountability from state and federal lawmakers. While this type of assessment is certainly a worthy goal, the paper contends that researchers in this area have been on a fool's errand: it is simply impossible to use these methodologies to claim that an individual institution is over- or under-performing in any meaningful way.

To understand the reason behind this conclusion one must first understand the two main purposes of estimating statistical models in the social sciences. By far the most common purpose has been hypothesis testing: does an individual variable have an impact on the phenomenon under study? The second purpose has been prediction: how does the phenomenon under study change for an individual observation given changes in the predictor variables? The results of a hypothesis test in a good model are usually stable given small changes in the data because standard hypothesis tests generally yield a yes/no answer based on the size of the coefficient and other information about the sample. The individual predictions for each observation, however,

are not necessarily stable, since the predicted value is not a binary outcome but is instead an actual number. Changes in the estimated coefficients that would not affect the results of a hypothesis test may have large effects on the predicted values for an individual observation. Herein lies the flaw in these graduation rate studies: small changes in sample selection, variable definition and model specification can yield large changes in predictions.

In addition, these studies fail to take into account the nature of the predicted values taken from the regression equations that are used to calculate the predicted graduation rates. These values are an econometric forecast (Kennedy 1993 p. 268) from an error-based statistical model, and as such contain error themselves. Confidence intervals should be reported for these forecasts to take this factor into account (similarly, public opinion polls reported on the evening news also report confidence intervals in the form of "60% of the American people support policy X, plus or minus three percentage points"). As will be seen, these confidence intervals often bracket the actual graduation rates for many institutions, yielding the conclusion that the predicted rates do not significantly differ from the actual rates. Yet USN and other researchers report these institutions as over- or under-performers, while the models themselves indicate they are performing as expected.

The data used in this study are very similar to the data used in the most recent USN rankings for national universities (see Smith 1998 p. 37 for details about their sample; see the Appendix for a description of the differences between the two datasets and data sources). The remainder of the paper assesses the robustness of the graduation rate performance approach by examining four potential problem areas:

---

[2] The main differences between these studies have been the number and type of independent variables used to estimate predicted graduation rates and the time-to-degree used to calculate the dependent variable. In addition, the Astin and Howard et. al. studies use individual level data. The critiques in this paper apply equally to these studies.

- Small changes in the sample due to missing data or changes in how the sample of national universities is defined.
- Seemingly irrelevant changes in variable definition.
- Different model specifications that take into account additional measures of student quality and institutional constraints.
- Confidence intervals for the predicted values.

Robustness of the predicted graduation rates

Sample selection

Although not discussed in the graduation rate literature, changes in the sample of universities may have an impact on how well an individual institution is estimated to perform. A robust measure of graduation rate performance should be immune to such changes – if small changes in the sample causes the measure to fluctuate, it becomes difficult to defend any conclusion of over or under performance. Two factors may cause the sample to change. First, there may be missing data for some variables or for entire institutions. This is a common occurrence in national studies attempting to collect data from numerous universities. Second, the institutions used in these studies are defined to be "national" universities, with the implication that universities can gain or lose national status over time and thus change the makeup of the sample.

A close examination of the latest USN rankings shows that several institutions have a 'N/A' reported for their predicted graduation rates. For example, Union Institute in Ohio does not have a predicted graduation rate reported in the 1999 rankings. An examination of the 1992 rankings reveal that SAT scores were not reported by USN for that year, so it is likely that USN

was unable to collect SAT scores for Union Institute's Fall 1991 cohort.[3] Such missing data can pose a problem in any analysis, with the result that institutions with missing data on one or more variables must be thrown out of the analysis.[4] As more variables are added to the model, the probability of having missing data for a variable increases. In the National Graduation Rate Study, 75 institutions were solicited for data. Only 52 institutions responded with usable data files, and of these 52 institutions eight were removed from the analysis due to missing data for some variables (Howard et. al. 1994 pp. 2-3).

A more serious problem is the definition of the sample. USN relies on the higher education classifications developed by the Carnegie Foundation (Carnegie Foundation for the Advancement of Teaching, 1994). Their classifications are based on the number of graduate degrees awarded, number of disciplines offered, and federal support awarded. While such classification systems may be laudable, they depend on somewhat arbitrary cutoffs for the measures of interest. Slight changes in cutoffs will change the makeup of the sample, and if the cutoffs are held constant institutions will drift into and out of the sample over time as their programs and federal support change. Other sample definitions are even more arbitrary. Astin simply uses baccalaureate-granting institutions that participated in a survey, as did Howard et. al. (Astin 1997, p. 648, Howard et. al. 1995 p.1).

A comparison of the national university samples from the 1992 and 1998 rankings, which list score and graduation data for the 1991 cohorts, is illustrative. Of the 204 national universities in the 1992 rankings, seven were dropped from the national university sample in 1998. Thirty-

---

[3] USN confirms that institutions are excluded based on a lack of historical data (Graham 1999). The reason behind some of the other N/A's, however, is a mystery. Duquesne University of Pennsylvania does not have a predicted graduation rate, yet their Fall 1991 SAT scores (1000) are reported in the 1992 rankings and their FY1992 expenditures ($7,749) are reported in the 1993 rankings.

[4] An alternative approach is to impute the missing values based on the remainder of the sample (Little and Rubin 1987). King et. al. (1998) show the deleterious impact of case deletion due to missing values and offer a simple method for imputing missing values.

one universities not in the 1992 sample were added to the 1998 sample, resulting in a total of 228 national universities in 1998.

The impact of missing data and changes in sample definition can be simulated in two ways. First, institutions can be randomly removed from the sample to estimate the impact of potential missing data. Second, institutions from the bottom quartile of the rankings can be removed to simulate reclassification of institutions as national universities. Institutions in the bottom quartile of the rankings are chosen because those in the top three quartiles are less likely to experience such extreme changes in major and degree programs as to cause their reclassification.

The graduation equation used by USN is presented in column 1 of Table 1. The six-year graduation rate for the 1991 new freshman cohort in 198 institutions is regressed on the average SAT score for the cohort[5] and the average expenditures per student by the institution (defined as the average spending per FTE student on instruction, research, student services and related educational expenditures). Ten institutions were randomly removed from both the entire sample and the bottom quartile and the graduation equation re-estimated. This process was then repeated an additional four times to estimate the graduation equation on ten different samples. The results are presented in Table 2, which shows the distribution of the *change* in predicted graduation rate from the full sample model for each of the trials for both simulations.

Simply removing ten random institutions causes the predicted graduation rates to fluctuate plus or minus one percentage point. The changes occur on a sizable portion of the sample: in some trials fully15-20% of the institutions changed predicted graduation rates. The sample change simulation produced similar results. Randomly removing ten institutions from the

group in the bottom quartile of the rankings also causes the rates to fluctuate plus or minus one percentage points (and in one trial, two percentage points). Moreover, the number of institutions affected by the change is much larger, affecting around 25-50% of the sample.

From the results we can conclude that both missing data and different sample definitions will affect the predicted graduation rates from these models, even though the amount of change may not seem remarkable. But two points should be kept in mind. First, as researchers it easy to dismiss such small changes as insignificant. But from the perspective of an individual institution, every percentage point counts: a two-point drop in their predicted graduation rate may seem very large indeed. Second and more importantly, the simulations reveal that the data used cannot be considered the population of national universities; instead, the data must be viewed as a *sample* of all national universities. This seemingly innocuous distinction has a very important implication as to how we treat the predicted rates and will be discussed below.

Variable definition

In addition to sample definition, there are several ways to define the explanatory variables in the graduation rate model. Should the quality of the cohort be expressed as the mean or the median SAT score? Either method is justifiable, yet it is likely that the two methods would yield different predicted graduation rates. Unfortunately the data are not available to test this possibility.

The expenditure data, however, is available for testing. In the latest rankings USN uses the average spending per FTE student on instruction, research, student services and related educational expenditures as their measure of financial resources available to the institution

[5] The SAT scores reported by USN are actually averages and midpoints if the institution only reported 25th and 75th percentile scores. ACT scores were converted to SAT scores using a College Board concordance table (Marco et. al.

(Smith 1998, p. 35). This variable is averaged over FY 1992-1995 (the only years for which IPEDS data are available for the 1991 cohort) and is used as the expenditure variable in the model listed in column 1 of Table 1. In the 1992 rankings USN used a slightly different measure of financial resources: not only the sum of educational expenditures but also all other spending, including such areas as research, scholarships and operations (Morse 1993, p.107). The graduation model using this second formulation is shown in column 2 of Table 1.

The equations are quite similar: the coefficients change slightly, and the predictive ability of the equations as measured by the adjusted R-square and the standard error of the regression is the same. The one difference is that the spending variable in the second model is now significant at the .05 level. But since the error levels are similar for both variables (p=.08 and p=.02), this difference is not as great as it might seem. The bottom portion of Table 1 shows the distribution of the difference between an individual institution's actual and predicted graduation rates. Again, the two models seem very much alike. The results would appear to meet expectations about the impact of slightly changing the definition of an explanatory variable: the two measures of spending are highly correlated (r=.97) and thus the statistical results are similar.

Yet the predicted graduation rates do differ. Table 3 shows the changes in predicted graduation rates when the total expenditures spending variable from the second model is used instead of the reduced spending variable in the first model. The predicted graduation rates for over half the sample change, and for seven institutions the rates changes by plus or minus two percentage points.

The results in Table 3 point to another problem with studies of this type. Seemingly irrelevant changes in variable definition may have little impact on the statistical results, yet still cause substantial changes in the predicted values of the dependent variable. Careful thought must

---

1992, p. 10).

be put into how variables are defined before one can reach the conclusion that individual institutions are over- or under-performing.

Model specification

More serious than sample selection or variable definition is the specification of the graduation performance model. An econometric model is said to be correctly specified when it describes or represents the underlying process of interest. Model specification can be a dicey business since reasonable researchers can often differ as to whether a model has been correctly specified. Yet improper specification, as in the case where a theoretically relevant variable has been excluded from the equation, can yield biased coefficients in a regression model. And since the coefficients are used to calculate a predicted graduation rate for individual institutions, poor model specification is not something that can be ignored.

With the graduation rate model we can attempt to assess model specification by asking two simple questions. Given that the model tries to measure student inputs to an institution as well as the constraints faced by an institution, does the model fully capture all of the relevant characteristics of the inputs, i.e. the incoming student cohort? And does the model capture all, or even most, of the relevant constraints faced by an individual institution?

In the case of the USN model the answers are clearly no. Certainly other aspects of the incoming student cohort besides their SAT/ACT scores will affect their six-year graduation rate. And universities face other constraints on their behavior besides the amount of money they are able to spend. The remainder of this section describes other available variables that theoretically could be included in the model and assesses the impact of their inclusion on the predicted graduation rates.

Additional measures of inputs and constraints

Average SAT/ACT scores are included in the graduation models to measure an institution's "input": Harvard attracts much more talented students than the University of Maryland, and this difference should be taken into account before either institution is judged on how well it graduates its students. But focusing only on standardized tests forces us to narrowly define academic credentials while also ignoring other aspects of the incoming student body. I consider three: the proportion of minorities in the cohort, the proportion of females, and the proportion of the student body over 25 years of age.

To a certain extent academic credentials will be reflected in standardized scores, but standardized scores cannot capture all aspects of how well a child has been educated. Having access to college preparatory courses, for example, will affect how many credits a student brings to college and how soon they will graduate. In addition, family income will also affect the ability of students to stay in school and graduate. Given that some institutions recruit more students from disadvantaged backgrounds than others, this must be taken into account when assessing performance.

Unfortunately such data are difficult to find at the institutional level. As a proxy for the academic and financial background of the cohort I use the proportion of the Fall 1991 cohort that is African-American (Sanford 1982). Access to primary education in this country is not equal across racial groups, and income disparities between blacks and whites are well known. Including this variable in the model helps take into account the fact that institutions recruit very different student bodies.

Differences in academic performance between males and females are also well known (Brower 1992, Hoskins et. al. 1997, Ronco 1996, Sanford 1982). Many universities are currently

experiencing trouble recruiting male students, and some institutions do a better job of recruiting male students than others. Just as an institution should not be penalized for failing to recruit students with perfect SAT scores, they should also not be penalized because their incoming student cohort is not evenly split between genders. I control for this by including the proportion of the incoming cohort that is female. Race/ethnicity and gender are standard control variables in models of retention and should be in any model of graduation rates. Both the gender and ethnicity variables come from the IPEDS datasets.

Finally, the average age of the incoming cohort is included in the expanded model (Breaugh and Mann 1981, Brower 1992, Hoskins et. al. 1997). Urban commuter schools, for example, tend to serve older student populations. Older students face a far different series of challenges than students fresh out of high school, such as financing issues and family responsibilities. Unfortunately the average age of the cohort is not collected by IPEDS or any other data source that I could find. As a proxy I include the proportion of the undergraduate student body over 25 years of age in Fall 1991 taken from Peterson's college guide. The average age of the new student cohort should be highly correlated with the average age of the undergraduate student body, and any bias resulting from differences between the two should be more than offset by the reduction in bias by including a relevant independent variable.

I now consider additional ways of measuring constraints. Certainly the amount of money an institution is able to spend on its students is an important constraint on its ability to graduate them in a timely manner. But institutions just as certainly face other constraints that affect their ability to graduate students that should also be taken into account. I briefly discuss three such possible constraints: whether an institution is public or private, the total enrollment of the institution and whether the institution has a religious affiliation.

Public institutions face a far different set of constraints than private institutions. They are usually overseen by one or more state agencies that regulate virtually all of the university's internal policies, from setting tuition rates and raising funds for capital improvements to determining faculty workloads and hiring and firing employees. Private institutions have much more freedom to set policies that may affect student performance, and this freedom should be taken into account in these models.

Total undergraduate enrollment is also a constraint that should have an impact on graduation rates (Huffman and Schneiderman 1997, Knox et. al. 1992, Pascarella 1985, Pascarella et. al. 1988). While in theory universities have the ability to enlarge or reduce enrollments as they please, in reality the infrastructure that has developed to support a certain size student body cannot be easily discarded. Many faculty members have tenure and cannot be fired, and union agreements may preclude massive reductions in staff. In addition, student body enrollments are often fixed by the state legislature for public universities and cannot be easily changed. So the total undergraduate enrollment at an institution should be viewed more as a constraint than a policy or program.

While student body size may affect graduation rates, the direction is not clear. Smaller institutions are generally equated with smaller class size and more student-faculty interactions. Given the emphasis on integration within the university community in most of the retention literature (e.g. Bean 1990, Tinto 1987), it seems likely that students would find it easier to become integrated in a university with a small undergraduate enrollment, thus increasing graduation rates. Large student enrollments, however, may work in the opposite direction due to returns to scale. With larger student enrollments it becomes cheaper to provide expensive

infrastructure such as research laboratories, recreation facilities or intensive student advising. Such infrastructure should also positively affect student behavior and thus graduation rates.

Finally, an institution's religious affiliation may also affect graduation rates (Mueller 1980). Universities with a religious affiliation are likely to provide a different atmosphere for their students than most institutions, as well as recruit different student bodies. Religious affiliations are usually historical in nature and cannot be considered something which universities can change as they please: in other words, a constraint.

In the expanded graduation model I include two dummy variables indicating whether an institution is public or has a religious affiliation, as well as the total undergraduate enrollment in Fall 1991. All data are from the IPEDS datasets.

Results

One problem associated with the inclusion of additional explanatory variables is missing data. For example, Peterson's did not report age data for several institutions. 19 institutions were thus dropped from the analysis due to missing data. The base graduation model was re-estimated for these institutions and the results presented in column 3 of Table 1. The estimates for the expanded model are presented in column 4. Several differences between the two equations are noteworthy.

First, the impact of SAT and spending on graduation rates change – the impact of SAT scores is reduced while the impact of spending increases. Second, three of the six additional variables are significantly related to graduation rates. All else being equal, institutions with large undergraduate enrollments, large proportions of females and younger students had higher graduation rates. Third, the expanded model does a better job of predicting graduation rates. The

standard error of the regression is almost two points lower in the expanded model. (The standard error of the regression or SEE is an estimate of the standard deviation of the forecast errors and is a vastly superior measure of model fit compared to the R^2 - see Achen 1982, pp. 58-68). In other words, the base model can on average predict graduation rates plus or minus 9.9 percentage points, but the expanded model's predictions fall within a smaller band, plus or minus 7.8 percentage points. This can be seen in the spread of the difference between actual and predicted graduation rates at the bottom of Table 1. Over 26% of the institutions in the reduced sample have predicted graduation rates ten percentage points greater or less than their actual graduation rates. Using a more properly specified model this proportion drops to slightly more than 12%.

All of these differences illustrate how serious model specification is when attempting to predict graduation rates. Additional explanatory variables did not change hypotheses concerning the impact of SAT scores on graduation rates: SAT scores have a significant impact in both models. But the additional variables did affect the predicted graduation rates: the proportion of institutions with extreme performance differences dropped.

This point becomes clear when the data are presented in a different manner. Table 4 classifies the 179 institutions by whether they were under-performers (their predicted rates were higher than their actual rates), no difference between the two rates, or over-performers (their predicted rates were lower than their actual rates). If model specification made little difference in the predicted rates, we would expect to see all of the institutions fall in the bolded cells in a diagonal line. This is clearly not the case. By adopting a more complex view of the inputs and constraints faced by a university, we can cause 30 institutions to suddenly change from under-performers to over-performers, or vice-versa (the changes here are not simply institutions

moving from +1 to –1 in terms of performance; many of the institutions experienced swings of over 10 percentage points).

Unfortunately model specification is one of the thorniest issues in econometric analysis. While most researchers might agree with the specification of the model in column 4 of Table 1, they might not agree with the specification in column 5. Two independent variables have been added: housing available to undergraduates, measured as the ratio of housing units available to undergraduates divided by the total undergraduate enrollment, and annual in-state tuition and fees (both as of the Fall 1991 semester). Both variables can be considered programs and policies set by the university. If housing is in demand, the university can simply build more housing. And if tuition is too high or low, the university is free to change it as well. In this view both variables should be excluded from any graduation model because on-campus housing and tuition are known to affect student behavior. Yet both these variables can also be considered constraints. Many universities may lack the space for additional dormitories, for example, or if public lack the power to change tuition rates. In this view they should be included in the model as constraints.

The SEE reveals that this model does an even better job of predicting graduation rates, and the distribution of over/under performance compared to the results in column 4 indicate that the predicted graduation rates fluctuate. Which model should be used? If both models can be justified on theoretical grounds, how reliable are the predicted graduation rates if they change with simple changes in model specification? This illustrates one of the major problems with research in this area. If simple changes in model specification can cause changes in predicted graduation rates, it becomes difficult to defend the practice of listing institutions by their

supposed graduation performance. How well an institution does depends not only on their programs and policies; performance also depends on the whims of the researcher.

Confidence intervals

This section considers the fourth criticism of the graduation rate models: the treatment of the predicted values from the regression equation. When estimating a regression model researchers report confidence intervals for the coefficients for each independent variable, usually in the form of standard errors or t-statistics. They do so because the data analyzed are usually from a sample of a larger population. Because of sampling variation we cannot be certain that the relationships found within the sample mirror those in the population, and confidence intervals allow us to assess the probability that the relationships in the sample would also be found in the population.

If the data used in the analysis consist of a population then confidence intervals are unnecessary, because there is no possibility that the relationships found do not exist in the population. Researchers in the graduation rate performance area implicitly use this assumption when reporting predicted graduation rates without confidence intervals. Unfortunately this assumption is unwarranted for two reasons.

First, we can never be certain that we have obtained data for the population of national universities. Consider the Carnegie classifications. If cutoffs based on number of programs and research dollars are used to define the population, how can we be sure we are not excluding universities that might be considered "national" from a theoretical point of view? And if the cutoffs are only redefined every few years, in the years between reclassifications some institutions will become national according to the definition, but will not be included in the

population until the next reclassification. If the data are collected by survey, the problem becomes even worse. If only one institution does not respond, then the data used for the analysis automatically becomes a sample. Examination of the Astin and Kroc et. al. analyses reveals non-response to be a serious problem, as well as deletion of institutions due to missing data. Table 2 indicates how much the results can vary when missing data is a problem.

Second, suppose we are somehow able to obtain data for the population of national universities. Undoubtedly mistakes have been made in the data collection process. For the data to reach its final form, programmers at individual institutions must run numerous computer programs; the data must be transcribed to the survey form; once collected, the surveys must be entered into a database for analysis. Errors can and will occur along every step, and if the process was repeated several times the resulting datasets would all differ in some small way. So even if data are obtained from a population of institutions, the data must be viewed as a sample of the "true" data and not the actual correct data itself.

Understanding that the predicted graduation rates have a random component has serious implications for what we can say about institutional performance. Error is introduced into the predicted graduation rates in several ways. First, even if the estimated coefficients (or relationships between the independent variables and graduation rates) in the sample are exactly the same in the population, the random nature of the error term in the regression model ensures that the predicted values will differ from the true predicted values. Second, it is unlikely that the estimates of the coefficients will exactly equal those in the population, introducing further error. Third, if the model if misspecified and does not represent the "true" model of graduation rate performance the predicted values will contain additional error (Pindyck and Rubinfeld 1981, p.205).

Such error must be taken into account when comparing predicted to actual graduation rates. The predicted rate may be higher or lower than the actual rate not because of institutional performance but simply because of random error. The calculation of a confidence or prediction interval for an individual forecast has a similar interpretation as the standard error of the regression: the predicted value is expected to fall within plus or minus a certain number of percentage points. If the actual graduation rate falls within this bracket, then we must conclude that there is no statistically significant difference between the forecasted graduation rate and the actual graduation rate for an institution.

Prediction intervals were calculated for both the full sample using the base USN model in column 1 of Table 1 and the reduced sample using the expanded model in column 4 (see Table 5). Only 5% of the institutions have predicted graduation rates that lie outside their prediction interval using a 95% confidence interval. Using the less rigorous 90% confidence interval the percentage only increases to 7-8%. This is a diametrically opposite conclusion compared to the 94% of institutions reported in the USN 1998 rankings as having differences between their predicted and actual rates.

Why do so few institutions have predicted graduation rates that significantly differ from their actual rates? Given the range of predicted values reported in the bottom of Table 1 and standard errors of the regression (in the +/- 8 to 10 percentage point range), one would expect many actual graduation rates to fall outside the prediction intervals. But rather than estimating a confidence interval for all observed values of the dependent variable, as in the case of the SEE, for individual institutions we are estimating an interval for only one observation. Such intervals are much less precise and in practice will be more than twice as large as the SEE (see the discussion in Pindyck and Rubinfeld 1981, pp. 206-211).

Taking into account the random nature of the data results in a very different conclusion about institutional performance. USN and other researchers regularly publish findings demonstrating that many institutions are radically over- or under-performing in terms of their ability to graduate their students. The finding here indicates the opposite: the vast majority of institutions are performing as expected. In retrospect such a conclusion is not surprising. Although universities operate in a far from perfectly competitive market, they still must compete for and retain students in order to survive as institutions. Such gross under-performance by institutions as reported by USN and other researchers is simply not credible.

Conclusion

The data presented here raise serious questions about the graduation performance enterprise embarked upon by many analysts. While attempts to hold institutions accountable and assess their performance are laudable, research in this area must be above reproach. The results of these models are taken seriously by institutions and can have quite an impact in the real world.[6] While publication of simple graduation rates has been charged as misleading (Astin 1990), the results indicate that "value-added" models that attempt to take into account an institution's inputs are as much, if not more, misleading.

This analysis demonstrates that the models used to assess graduation rate performance are highly sensitive to sample and variable definition and model specification. Given how predicted graduation rates fluctuate depending on what model is used and how its variables are defined, it becomes impossible to defend conclusions such as, "apparently Alaska institutions do not provide academically supportive environments leading to graduation within six years [because

their actual rates are less than their predicted rates]" (Mortenson 1997, p. 4). Worse, proper statistical use of the predicted graduation rates reveals that research in this area is very much a case of the Emperor's new clothes: 95% of the institutions in this study have predicted graduation rates that do not significantly differ from their actual graduation rates.

Recommendations for future research

Not surprisingly, my recommendation would be to cease this enterprise entirely. Given that only a fraction of the institutions in this study show statistically significant differences between actual and predicted graduation rates, and given the questionability of even this conclusion due to possible problems with variable definition and model selection, the possibility of obtaining meaningful results appears almost nonexistent. Since the demand for accountability indicators will certainly increase, researchers will continue to work in this area, so I offer some recommendations for improving this body of research.

- **Be very careful about how you define the sample and variables. Do not let the data drive your analysis.**

The Astin and Kroc et. al. projects are perfect examples of how not to conduct this type of research. The arbitrary nature of their samples calls into question all of their results, and researchers would be wise to adopt the USN approach of *first* defining the theoretically relevant population to study and *then* collecting the data. Greater efforts must be made to minimize case deletion due to missing data.

- **Spend time developing a sound theoretical model of how inputs and constraints affect student behavior. Specify your model accordingly**.

---

[6] See Mufson (1999) for a description of how George Washington University's business school spent $1.5 million and radically restructured its program solely to improve its position in USN's rankings. But Hossler and Foley

Here researchers must avoid the pitfalls of both a minimally specified model as typified by the USN approach as well as an overspecified model as typified by Kroc. et. al. Careful thought must be put into the selection of variables for the model. Proper modeling procedures must also be followed. Removal of variables from the model due to statistical insignificance, for example (e.g. Astin 1996, p. 17), should be avoided (Achen 1982, pp. 51-68).

- **Understand the nature of the predictions from your model. Report confidence intervals for predictions.**

Forecasts from an error-based model *must* take this error into account. As the analysis has demonstrated, taking such error into account can lead to drastically different conclusions.

(1995) argue that the rankings have little impact on student choice.

Appendix – Data Description and Sources

The USN dataset consists of 218 of the 228 institutions in their national university sample (ten institutions are listed as 'N/A' for their predicted graduation rate – I assume these institutions were excluded from the data they analyzed). The data analyzed in this paper contain only 198 institutions.

Two of the 218 institutions, the University of Memphis and the University of Alabama at Tuscaloosa, do not appear in the 1992 IPEDS enrollment data and were excluded from the analysis. In addition, SAT/ACT data for many institutions were not listed in the 1992 rankings. After querying the institutions I was able to fill in the data for some schools, leaving 198 out of the original 218 institutions in my dataset.

Table A-1 lists some information on the predicted graduation rates for both samples. The predicted graduation rates statistics reported for USN are calculated on the rates published in the latest rankings. Following USN, predicted values greater than 100 were reset to … Source? Despite the lack of an exact match between the two datasets, they do appear quite similar.

Sources for the data used in the analysis are given in Table A-3.

References

Achen, C. H. *Interpreting and Using Regression*. Beverly Hills: Sage Publications, 1982.

Astin, A.W. "Retention-Rate Data Mislead Student Consumers." *Chronicle of Higher Education* (Nov. 21, 1990), p. B2.

Astin, A. W. "How 'Good' is Your Institution's Retention Rate?" *Research in Higher Education*, 38 (1997), 647-658.

Bean, J. "Why students leave: insights from research." In D. Hossler et. al. *The Strategic Management of College Enrollments*. San Francisco: Jossey-Bass Publishers, 1990.

Breaugh, J. A. and R. B. Mann. "The Utility of Discriminant Analysis for Predicting Graduation from a Master of Business Administration Program." *Educational and Psychological Measurement*, 41 (1981), 495-501.

Brower, A.M. "The Second-Half of Student Integration: The Effects of Life Task Predominance on Student Persistence." *Journal of Higher Education*, 63 (1992) 441-462.

Dilts, S. (ed.). *Peterson's Guide to Four-Year Colleges.* Princeton: Peterson's Guides, 1993.

Elfin, M. and B. Brophy. "America's Best Colleges," *U.S. News and World Report*, 113 (September 28, 1992), 96-127.

Geraghty, M. "Student Opposition to 'U.S. News' College Rankings Mount." *Chronicle of Higher Education* (November 15, 1996), p. A50.

Hoskins, S.L., S.E. Newstead and I. Dennis. "Degree Performance as a Function of Age, Gender, Prior Qualifications and Discipline Studied." *Assessment & Evaluation in Higher Education*, 22 (1997), 317-328.

Hossler, D. and E. M. Foley. "Reducing the Noise in the College Choice Process: The Use of College Guidebooks and Ratings." In Walleri, R. and M. Moss, eds. *Evaulating and Responding to College Guidebooks and Rankings*. San Francisco: Jossey-Bass Publishers, 1995.

Howard, R., R. Kroc and D. Woodward. *National Graduation Rate Study 1994: Report 1*. 1994.

Huffman, J.P. and S. Schneiderman. "Size Matters: The Effect of Institutional Size on Graduation Rates." Paper presented at the Association for Institutional Research meeting, Orlando, 1997.

Kennedy, P. *A Guide to Econometrics*. Cambridge: The MIT Press, 1993.

King, G., J. Honaker, A. Joseph and K. Scheve. "Listwise Deletion is Evil: What to Do About Missing Data in Political Science." Manuscript, Department of Government, Harvard University. 1998. Paper available online at http://polmeth.calpoly.edu/working98.alpha.html.

Knox, W. E., P. Lindsay and M. N. Kolb. "Higher Education, College Characteristics and Student Experiences: Long-Term Effects on Educational Satisfactions and Perceptions." *Journal of Higher Education*, 63 (1992), 303-328.

Kroc, R., D. Woodard, R. Howard and P. Hull. "Predicting Graduation Rates: a Study of Land Grant, Research I and AAU Universities." Paper presented at the Association for Institutional Research meeting, Boston, 1995.

Kroc, R., D. Woodard, R. Howard, P. Hull and D. Woodward. "Graduation Rates: Do Students' Academic Program Choices Make a Difference?" Paper presented at the Association for Institutional Research meeting, Orlando, 1997.

Machung, A. "Playing the Rankings Game." Change (July/August 1998), 13-16.

Marco, G.L., A.A. Abdel-Fattah and P.A. Baron. *Methods Used to Establish Score Comparability on the Enhanced ACT Assessment and the SAT*. 1992. College Board Report No. 92-3. New York: College Entrance Examination Board.

Morse, R. J. "Methodology." *U.S. News and World Report*, vol. 115, no. 13 (October 4, 1993), 107.

Mortenson, T. "Actual versus Predicted Institutional Graduation Rates for 1100 Colleges and Universities," *Postsecondary Education Opportunity*, no. 58 (April 1997).

Mueller, C. "Evidence on the Relationship between Religion and Educational Attainment." *Journal of Higher Education*, 53 (1980) 140-152.

Mufson, S. "Rankings All-Important to GWU." *Washington Post*, March 14, 1999, A01.

Pascarella, E.T. "Students Affective Development within the College Environment." *Journal of Higher Education*, 56 (1985), 640-663.

Pascarella, E.T., C.A. Ethington and J.C. Smart. "The Influence of College on Humanitarian/Civic Involvement Values." *Journal of Higher Education*, 59 (1988), 412-437.

Pindyck, R. S. and D. L. Rubinfeld. *Econometric Models and Economic Forecasts*. New York: McGraw-Hill Book Company.

Ronco, S.L. "How Enrollment Ends: Analyzing the Correlates of Student Graduation, Transfer and Dropout with a Competing Risks Model." *AIR Professional File*, 61 (1996).

Sanford, T. "Predicting College Graduation for Black and White Freshman Applicants." *College and University* 57(1982) 265-278.

Schmitz, C. "Assessing the Validity of Higher Education Indicators," *Journal of Higher Education*, 64 (1993), 503-521.

Shea, C. "Annual College Rankings Roil but Also Gain in Influence." *Chronicle of Higher Education* (September 22, 1995), p. A53.

Smith, S. (ed.). *America's Best Colleges, 1999*. Washington, D.C.: U.S. News and World Report, 1998.

Tinto, V. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago: University of Chicago Press, 1987.

Webster, D. (1992). Rankings of undergraduate education in "U.S. News and World Report" and "Money": are they any good? *Change* 24(2):18-31.

Table 1. Models of Six-Year Graduation Rates, Fall 1991 Cohorts

	Full sample		Reduced sample		
	1	2	3	4	5
Intercept	-83.394**	-88.879**	-83.469**	-92.547**	-98.893**
	(14.095)	(13.192)	(14.692)	(16.523)	(15.771)
SAT	0.105**	0.102**	0.103**	0.084**	0.082**
	(.008)	(.007)	(.008)	(.008)	(007)
Logged spending I	3.556		3.792	5.623**	5.860**
	(2.029)		(2.127)	(1.943)	(1.910)
Logged spending II		4.231*			
		(1.801)			
Public institution				-0.035	0.751
				(2.227)	(3.897)
Total enrollment				0.138*	0.237**
				(.065)	(.065)
Religious affiliation				2.135	4.165
				(2.578)	(2.596)
% African-American				0.061	0.009
				(.078)	(.075)
% female				0.307**	0.257**
				(.068)	(.065)
% of undergraduates > 25				-0.367**	-0.277**
				(.045)	(.048)
Housing availability					16.493**
					(3.580)
Annual tuition and fees					-0.186
					(.339)
F statistic	253.6	258.0	213.5	99.6	90.8
adjusted R^2	0.72	0.72	0.70	0.82	0.83
SEE	9.8	9.8	9.9	7.8	7.4
N	198	198	179	179	179
Distribution of over/under performance[a]					
21% and over	0.5	0.5	0.6	1.1	0.6
11% to 20%	12.6	13.6	12.3	5.0	7.3
1% to 10%	38.4	38.4	39.1	39.7	36.9
0%	4.0	4.5	3.9	7.3	9.5
-1% to –10%	32.3	30.8	30.7	40.8	41.3
-11% to –20%	9.6	9.6	10.6	5.6	3.9
-21% and under	2.5	2.5	2.8	0.6	0.6
	100.0	100.0	100.0	100.0	100.0

Note: * p<.05, ** p<.01.
[a]Defined as actual graduation rate minus predicted.

Table 2. Change in Predicted Graduation Rates Due to Changes in Sample

	Change in predicted graduation rate from full sample model	Trial				
		1	2	3	4	5
Missing data simulation (10 institutions randomly removed from full sample)	- 1	2	19	25	40	4
	0	168	167	157	147	159
	+ 1	18	2	6	1	25
Sample change simulation (10 institutions randomly removed from bottom quartile of sample)	- 2	0	0	10	0	0
	- 1	45	43	85	47	20
	0	137	145	69	128	132
	+ 1	6	0	22	13	36
	+ 2	0	0	2	0	0

Note: cell entries are numbers of institutions.

Table 3. Change in Predicted Graduation Rates Due to Different Definitions of Spending

Change in predicted graduation rate using total expenditure variable	N	%
- 2	5	2.5
- 1	48	24.2
0	86	43.4
+ 1	57	28.8
+ 2	2	1.0

Table 4. Changes in Over- and Under-Performance, Base and Expanded Models

		Expanded model			
	Performance:	*Under*	*No change*	*Over*	TOTAL
Base model	*Under*	**63**	3	13	79
		35.2%	1.7%	7.3%	44.1%
	No change	4	**1**	2	7
		2.2%	**0.6%**	1.1%	3.9%
	Over	17	9	**67**	93
		9.5%	5.0%	**37.4%**	52.0%
	TOTAL	84	13	82	179
		46.9%	7.3%	45.8%	100.0%

Note: institutions classified using predicted rates from equations in columns 3 and 4 of Table 1.
Under is actual < predicted, no change is actual = predicted and over is actual > predicted.

Table 5. Percentage of Institutions Whose Predicted Graduation Rates Significantly Differ from their Actual Graduation Rates

	Using 95% confidence intervals		Using 90% confidence intervals	
	N	%	N	%
Base model[a]	9	4.5%	16	8.1%
Expanded model[b]	9	5.0%	13	7.3%

[a]Using equation in column 1 of Table 1.
[b]Using equation in column 4 of Table 1.

Figure 1. Actual and Predicted Six-Year Graduation Rates from USN *Best Colleges* 1999

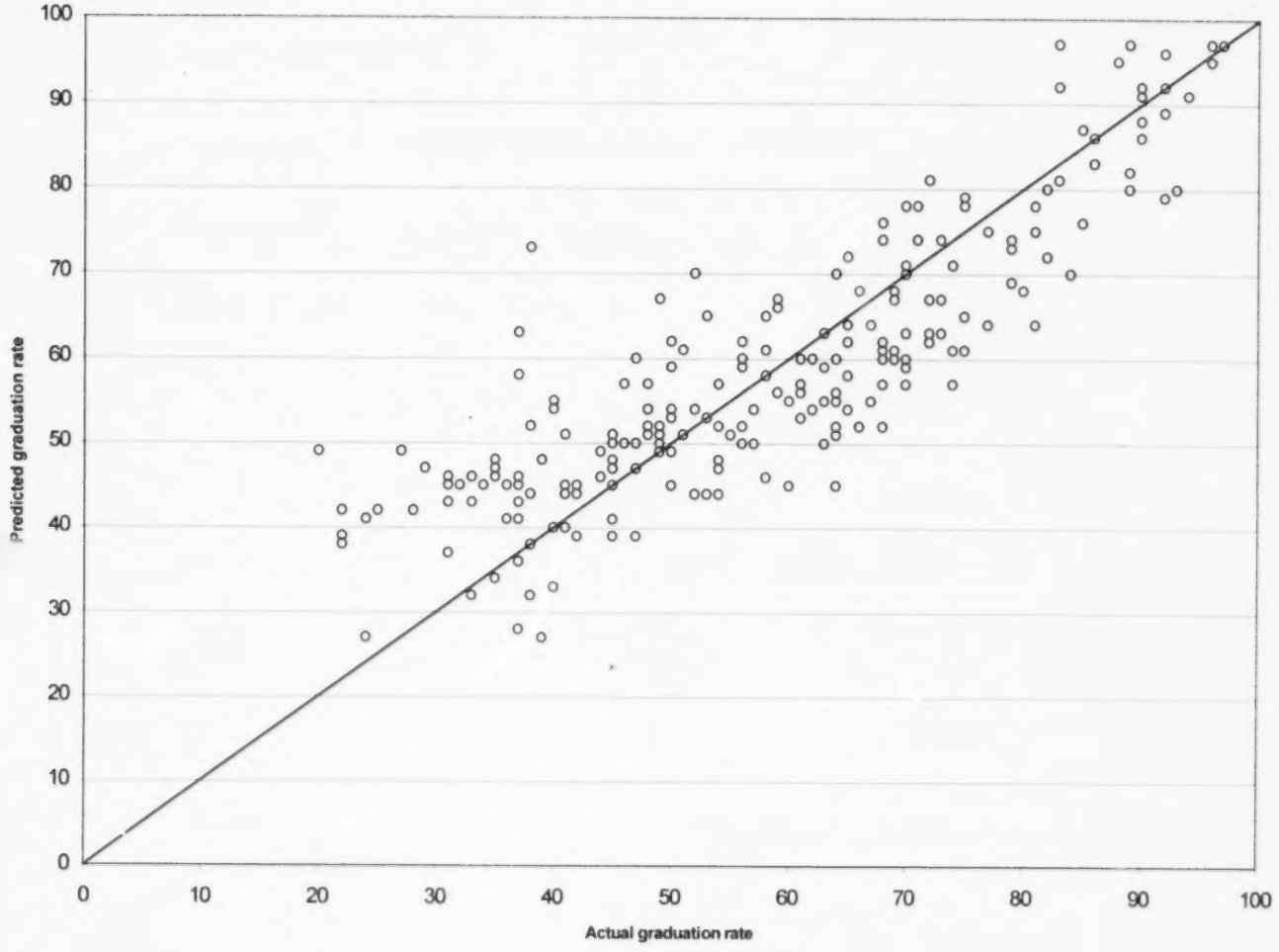

Table A-1. Comparison of Predicted Graduation Rates

Variable	Statistic	USN[a] (N = 218)	Full sample[b] (N = 198)	Reduced sample[b] (N = 179)
Predicted graduation rate	Mean	59.0	59.7	59.3
	Minimum	27	26	26
	Maximum	97	97	97
	Standard deviation	15.5	15.6	15.3
	Correlation with USN	1.00	.97	.97

[a]Taken from Smith 1998.
[b]Based on equations in columns 1 and 3 of Table 1.

Table A-2. Data Sources

Variable	Source
Six-year graduation rate	Smith 1998
SAT/ACT scores	Elfin and Brophy 1992 or the institution
Fiscal year expenditures	IPEDS
Gender and minority data	IPEDS
Student body totals	IPEDS
Institution type	IPEDS
Religious affiliation	IPEDS
% of undergraduates 25 and over	Dilts 1993
Annual in-state tuition and fees	Dilts 1993
Number of housing spaces for undergraduates	Dilts 1993